

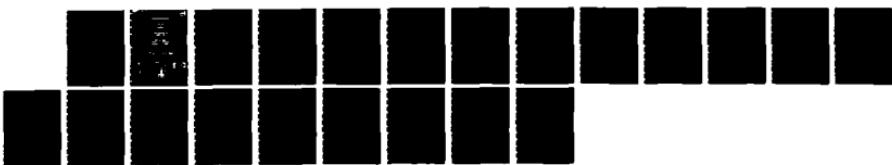
AD-A175 850

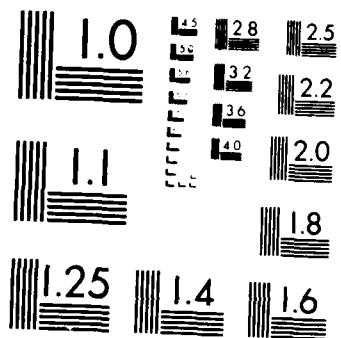
QUANTILES OF KAPLAN-MEIER ESTIMATOR(U) STANFORD UNIV CA 1/1
DEPT OF STATISTICS H DOSS ET AL 84 NOV 86 TR-382
N80014-86-K-0156

UNCLASSIFIED

F/G 12/1

NL





MRCOPY RESOLUTION TEST CHART
NATIONAL RESEARCH COUNCIL, CANADA

AD-A175 050

QUANTILES OF KAPLAN-MEIER ESTIMATOR

BY

HANI DOSS

and

STEVEN FREITAG

TECHNICAL REPORT NO. 382

NOVEMBER 4, 1986

PREPARED UNDER CONTRACT
N00014-86-K-0156 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. INTRODUCTION AND SUMMARY.

Let (X_i, Y_i) , $i = 1, \dots, n$ be independent identically distributed pairs of positive random variables with distribution functions $F(t) = P(X_i \leq t)$ and $G(t) = P(Y_i \leq t)$. Suppose that X_i and Y_i are independent for each i . In the random censorship model the X 's represent survival times and the Y 's represent censoring times; the observations consist of (Z_i, δ_i) , $i=1, \dots, n$, where $Z_i = \min(X_i, Y_i)$ and $\delta_i = I(X_i \leq Y_i)$.

The most commonly used estimate of F is the Kaplan-Meier estimator defined by

$$\hat{F}(t) = 1 - \prod_{Z(i) \leq t} \left(\frac{n-i}{n-i+1} \right)^{\delta(i)} \quad (1.1)$$

where $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)}$ denote the ordered values of Z_1, Z_2, \dots, Z_n , and $\delta_{(i)}$ is the δ corresponding to $Z_{(i)}$. Note that this version of the Kaplan-Meier estimator is strictly positive on $[Z_{(n)}, \infty)$ if the last observation is censored and is therefore slightly different from the version which is always 0 on $[Z_{(n)}, \infty)$, originally proposed by Kaplan and Meier (1958); see Wellner (1985) for a discussion.

An important problem in survival analysis is to estimate the p^{th} quantile of F defined by $\xi_p = \inf\{t: F(t) \geq p\}$. The natural estimate of ξ_p is

$$\hat{\xi}_p = \inf\{t: \hat{F}(t) \geq p\}. \quad (1.2)$$

The purpose of this paper is three-fold. First, in Section 2 we show how the theory of counting processes, martingales and stochastic integration can be used to obtain in a simple way the following result.

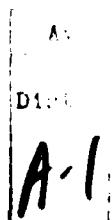
THEOREM 1. Suppose that the following conditions hold for $p \in (0,1)$.

- (i) F is continuous and $G(\xi_p) < 1$.
- (ii) F has a positive derivative at ξ_p .

Then

$$\hat{\xi}_p = \xi_p + \frac{p - \hat{F}(\xi_p)}{F'(\xi_p)} + R_{n,p}, \quad (1.3)$$

where $n^{-1} R_{n,p} \rightarrow 0$ in probability as $n \rightarrow \infty$.



Since $n^{\frac{1}{2}}(\hat{F}(\xi_p) - p)$ converges in distribution to a normal random variable with mean 0 and variance

$$(1-p)^2 \int_0^{\xi_p} \frac{dF(u)}{(1-G(u))^2(1-F(u))}, \quad (1.4)$$

(see, e.g. Gill, 1983) Theorem 1 immediately yields that $n^{\frac{1}{2}}(\hat{\xi}_p - \xi_p)$ is asymptotically normal with mean 0 and variance

$$\left(\frac{1-p}{F'(\xi_p)} \right)^2 \int_0^{\xi_p} \frac{dF(u)}{(1-G(u))^2(1-F(u))}. \quad (1.5)$$

We remark that (1.3) is neither new nor the strongest version of this result. Aly, Csörgő and Horváth (1985) obtained under more restrictive conditions than those given in Theorem 1, a uniform (in p) rate of convergence for $R_{n,p}$ of $O(n^{-\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log \log n)^{\frac{1}{4}})$, as $n \rightarrow \infty$ (see Proposition 3.1 of Section 3 in the present paper for a statement of this result). Aly et al. (1985) assume in addition to the conditions given in Theorem 1 that G is continuous and that F is twice differentiable at ξ_p in order to obtain for fixed p their rate of convergence for $R_{n,p}$. Therefore, Theorem 1 is of interest because of its simple proof and because its assumptions are less restrictive than those of Aly et al. (1985).

Our second purpose is to prove the asymptotic independence of the estimates of quantiles in the competing risks problem; this is done in Section 3. The competing risks problem arises in medical follow-up studies and in industrial life testing when the failure of an individual can be classified into one of m mutually exclusive classes, referred to as "risks"; see Gail (1975). Consider a sample of n individuals each of whom can incur a failure by any one of m risks. Let X_{ij} , $i = 1, \dots, n$, $j = 1, \dots, m$ denote the lifelength of individual i if risk j is the sole cause of failure. Let F_j denote the distribution of X_{ij} . Assume that X_{i1}, \dots, X_{im} are independent, and that the vectors (X_{i1}, \dots, X_{im}) $i = 1, \dots, n$ are independent. Since the life of individual i is terminated by the earliest occurring cause of failure, one observes only the sequence $\{(Z_i, \delta_{ij}), j=1, \dots, m, i=1, \dots, n\}$, where $Z_i = \min(X_{i1}, \dots, X_{im})$ and $\delta_{ij} = I(Z_i = X_{ij})$.

Thus, δ_{ij} indicates whether risk j was the cause of death ($\delta_{ij} = 1$) or not ($\delta_{ij} = 0$).

Note that for each j , we can estimate F_j by

$$\hat{F}_j(t) = 1 - \prod_{Z_{(i)} \leq t} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)j}}, \quad (1.6)$$

where $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)}$ are the ordered values of Z_1, Z_2, \dots, Z_n and $\delta_{(i)j}$ are the corresponding values of $\delta_{1j}, \delta_{2j}, \dots, \delta_{nj}$.

The p^{th} quantile of F_j may be estimated by

$$\hat{\xi}_{j,p} = \inf\{t: \hat{F}_j(t) \geq p\}. \quad (1.7)$$

We combine the Bahadur representation of quantiles given by Aly, et al. (1985) and a weak convergence result of Aalen (1976) to show in a simple argument that when suitably normalized, the vector $(\hat{\xi}_{1,p}, \dots, \hat{\xi}_{m,p})$ converges in distribution to a vector of independent Gaussian processes.

Finally, in Section 4 we use the ideas in the proof of Theorem 1 to study the quantile estimates that arise in the reliability model introduced by Doss, Freitag, and Proschan (1986).

2. A REPRESENTATION FOR QUANTILES OF THE KAPLAN-MEIER ESTIMATOR.

An account of the theory of counting processes, martingales, and stochastic integration used in this section is given in Chapter 18 of Lipster and Shirayev (1978) and in Chapter 2 of Gill (1980). A very accessible exposition of this theory is provided in Anderson and Borgan (1984). Throughout the paper we adopt the convention that $\frac{0}{0} = 0$; $x \wedge y$ denotes $\min(x, y)$, and if H is a distribution function then $\bar{H}(t) = 1 - H(t)$.

The following processes are needed to give $(\hat{F}(t) - F(t \wedge Z_{(n)})) / \hat{F}(t \wedge Z_{(n)})$ a stochastic integral representation which we will use to prove our results.

$$J(t) = I(Z_{(n)} \geq t) \quad (2.1)$$

$$V(t) = \sum_{i=1}^n I(Z_i \geq t) \quad (2.2)$$

$$N(t) = \sum_{i=1}^n I(Z_i \leq t, \delta_i = 1) \quad (2.3)$$

$$A(t) = \int_0^t \frac{V(s)}{\bar{F}(s)} dF(s) \quad (2.4)$$

$$M(t) = N(t) - A(t) \quad (2.5)$$

Our proof of Theorem 1 will require the following two results taken from the literature.

PROPOSITION 2.1 Suppose F is continuous. Then

$$(i) \quad \frac{\hat{F}(t) - F(t \wedge Z_{(n)})}{\hat{F}(t \wedge Z_{(n)})} = \int_{(0,t]} \frac{\hat{F}(s-) J(s)}{\hat{F}(s) V(s)} dM(s) \quad (2.6)$$

for $t \geq 0$, where $\hat{F}(s-) = \lim_{u \uparrow s} 1 - \hat{F}(s)$, and

$$(ii) \quad \left\{ \left[\left(n^{\frac{1}{2}} \int_{(0,t]} \frac{\hat{F}(s-) J(s)}{\hat{F}(s) V(s)} dM(s) \right)^2 - \int_{(0,t]} \left[n^{\frac{1}{2}} \frac{\hat{F}(s-) J(s)}{\hat{F}(s) V(s)} \right]^2 dA(s), F_t \right] \right\}; t \in [0,T]$$

is a martingale, where F_t is the completion of the σ -field generated by $\sigma(I(Z_i \leq s, \delta_i = 1), I(Z_i \leq s, \delta_i = 0); 1 \leq i \leq n, s \leq t)$ and $T > 0$ is such that $F(T) < 1$.

PROOF. The identity (2.6) is given in Lemma 3.2.1 Part (iv) of Gill (1980), and also in Theorem 3.1 of Aalen and Johansen (1978).

To prove Part (ii), we note that it follows from Corollary 3.1.1 of Gill (1980) that $\{(M(t), F_t); t \in [0,T]\}$ is a martingale with quadratic variation process $\langle M \rangle(t) = A(t)$. Let $f(s) = (\hat{F}(s-) J(s)) / (\hat{F}(s) V(s))$. Since f is predictable, the integral $\int_{(0,t]} f(s) dM(s)$ is a martingale with quadratic variation process $\int_{(0,t]} f^2(s) d\langle M \rangle(s)$, and Part (ii) follows (see p. 268 of Lipster and Shirayev, 1978 or p. 10 of Gill, 1980; cf. also the paragraph preceding (4.1.11) page 55 of Gill, 1980).

PROPOSITION 2.2 Let $\{V_n\}_{n=1}^\infty$ and $\{W_n\}_{n=1}^\infty$ be two sequences of random variables satisfying the following conditions.

(i) For all $\sigma > 0$ there is a λ (depending on σ) such that $P(|W_n| > \lambda) < \sigma$ (2.7)
for all n .

(ii) For all k and $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(V_n \leq k, W_n \geq k + \epsilon) = 0 \text{ and } \lim_{n \rightarrow \infty} P(V_n \geq k + \epsilon, W_n \leq k) = 0 \quad (2.8)$$

Then $V_n - W_n \rightarrow 0$ in probability as $n \rightarrow \infty$.

PROOF: See Ghosh (1971).

PROOF OF THEOREM 1. We note that the assumption $F'(\xi_p) > 0$ implies that ξ_p is the unique number satisfying $F(\xi_p) = p$. For any $t \geq 0$ and any n define

$$z_{t,n} = n^{\frac{1}{2}}[F(\xi_p + \frac{t}{n^{\frac{1}{2}}}) - \hat{F}(\xi_p + \frac{t}{n^{\frac{1}{2}}})]/F'(\xi_p), \text{ and}$$

$$w_n = n^{\frac{1}{2}}(p - \hat{F}(\xi_p))/F'(\xi_p).$$

Note that

$$\{n^{\frac{1}{2}}(\hat{\xi}_p - \xi_p) \leq t\} = \{p \leq \hat{F}(\xi_p + \frac{t}{n^{\frac{1}{2}}})\} = \{z_{t,n} \leq t_n\}, \quad (2.9)$$

where

$$t_n = n^{\frac{1}{2}}\{F(\xi_p + \frac{t}{n^{\frac{1}{2}}}) - p\}/F'(\xi_p) \rightarrow t \quad (2.10)$$

by definition of the derivative. In (2.9) the first equality follows from the definition of $\hat{\xi}_p$ given by (1.2). Suppose we can show for fixed $t \geq 0$ that

$$z_{t,n} - w_n \rightarrow 0 \text{ in probability.} \quad (2.11)$$

Letting $V_n = n^{\frac{1}{2}}(\hat{\xi}_p - \xi_p)$, it would then follow from (2.9) and (2.10) that (2.8) holds for the sequences $\{V_n\}$ and $\{W_n\}$. Since W_n is asymptotically normal (see e.g. Theorem 4.2.2 of Gill, 1980) condition (2.7) is satisfied, and the theorem follows from Proposition 2.2.

It remains to prove (2.11). Algebraic manipulations yield that

$$\begin{aligned} z_{t,n} - w_n &= \frac{n^{\frac{1}{2}}}{F'(\xi_p)} \left(F(\xi_p + \frac{t}{n^{\frac{1}{2}}}) - F(\xi_p) \right) \left(1 - \frac{\hat{F}(\xi_p)}{\bar{F}(\xi_p)} \right) \\ &\quad + \frac{n^{\frac{1}{2}}}{F'(\xi_p)} \bar{F}(\xi_p + \frac{t}{n^{\frac{1}{2}}}) \left(\frac{\hat{F}(\xi_p + \frac{t}{n^{\frac{1}{2}}})}{\bar{F}(\xi_p + \frac{t}{n^{\frac{1}{2}}})} - \frac{\hat{F}(\xi_p)}{\bar{F}(\xi_p)} \right). \end{aligned} \quad (2.12)$$

The first product term on the right side of (2.12) converges to 0 in probability by the consistency of the Kaplan-Meier estimate at ξ_p (see e.g. Földes, Rejtö, and Winter, 1980) and the definition of the derivative.

We now proceed to show that the second product term on the right side of (2.12) converges to 0 in probability. For fixed t , there exists N (depending on t) such that for $n \geq N$, $F(\xi_p + \frac{t}{n^{\frac{1}{2}}}) < 1$. Thus, for $n \geq N$, by Proposition 2.1, the second product term on the right side of (2.12) is equal to

$$\begin{aligned} & - \frac{n^{\frac{1}{2}}}{F'(\xi_p)} \bar{F}(\xi_p + \frac{t}{n^{\frac{1}{2}}}) \int_{(\xi_p, \xi_p + \frac{t}{n^{\frac{1}{2}}})} \frac{\hat{F}(s-) J(s)}{\bar{F}(s) V(s)} dM(s) \\ & + \left[\left(\frac{\hat{F}(\xi_p + \frac{t}{n^{\frac{1}{2}}})}{\bar{F}(\xi_p + \frac{t}{n^{\frac{1}{2}}})} - \frac{\hat{F}(\xi_p + \frac{t}{n^{\frac{1}{2}}})}{\bar{F}((\xi_p + \frac{t}{n^{\frac{1}{2}}) \wedge Z(n)})} \right) - \left(\frac{\hat{F}(\xi_p)}{\bar{F}(\xi_p)} - \frac{\hat{F}(\xi_p)}{\bar{F}(\xi_p \wedge Z(n))} \right) \right]. \end{aligned} \quad (2.13)$$

It is clear that under the assumptions of the theorem, the probability that the quantity inside the brackets in (2.13) is not 0 converges to 0 (exponentially fast) as $n \rightarrow \infty$.

We now show that the first term in (2.13) converges to 0 in probability. Let $\epsilon > 0$ be such that $F(\xi_p + \epsilon) < 1$, and $G(\xi_p + \epsilon) < 1$. By the strong uniform consistency of the Kaplan-Meier estimate (see Földes, Rejtö, and Winter, 1980),

$$\sup_{0 \leq s \leq \xi_p + \epsilon} \left| \frac{\hat{F}(s)}{\bar{F}(s)} - 1 \right| \rightarrow 0 \quad \text{a.s.} \quad (2.14)$$

It follows from the Glivenko-Cantelli Theorem and (2.14) that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \int_{(\xi_p, \xi_p + \frac{t}{n^{\frac{1}{2}}})} \left[n^{\frac{1}{2}} \frac{\hat{F}(s-) J(s)}{\bar{F}(s) V(s)} \right]^2 dA(s) \\ & \leq \int_{\xi_p}^{\xi_p + \epsilon} \frac{1}{\bar{H}(s) \bar{F}(s)} dF(s) \quad \text{a.s.}, \end{aligned} \quad (2.15)$$

where $\bar{H}(s) = P(Z_1 > s)$. Note that $\bar{H}(s)$ is bounded away from 0 on $(\xi_p, \xi_p + \epsilon]$.

Since ϵ was arbitrary, (2.15) implies that

$$\int_{(\xi_p, \xi_p + \frac{t}{n^{\frac{1}{2}}})} \left[n^{\frac{1}{2}} \frac{\hat{F}(s-) - F(s)}{\bar{F}(s)} \frac{J(s)}{V(s)} \right]^2 dA(s) \rightarrow 0 \quad \text{in probability.} \quad (2.16)$$

We now apply Lenglart's inequality (Lenglart, 1977; for a statement, see Theorem 2.4.2 of Gill, 1980) to the martingale of Proposition 2.1, Part (ii) to conclude that

$$\left(n^{\frac{1}{2}} \int_{(\xi_p, \xi_p + \frac{t}{n^{\frac{1}{2}}})} \frac{\hat{F}(s-) - F(s)}{\bar{F}(s)} \frac{J(s)}{V(s)} dM(s) \right)^2 \rightarrow 0 \quad \text{in probability.}$$

This completes the proof of the theorem. ■

The proof of Theorem 1 generalizes the method used by Ghosh (1971) to the case of censored data.

To estimate the expression for the asymptotic variance of $n^{\frac{1}{2}}(\hat{\xi}_p - \xi_p)$ given by (1.5), we may estimate the term $(1-p)^2/(F'(\xi_p))^2$ and the integral term separately. Under the conditions of Theorem 1, the proposition in Section 2 of Hall and Wellner (1980) implies that $n \sum_{\substack{i: Z(i) \leq \hat{\xi}_p \\ i < n}} \frac{\delta(i)}{(n-i+1)(n-i)}$ is a consistent estimate of the integral in (1.5).

If we assume in addition to the conditions of Theorem 1 that F is continuously differentiable in a neighborhood of ξ_p , then Theorem 4.1.2 of Ramlau-Hansen (1983) implies that $F'(\xi_p)/\bar{F}(\xi_p)$ may be consistently estimated by a suitable kernel type of estimator; one such estimator is described in Section 5 of Ramlau-Hansen (1983).

Note that since ξ_p is unknown, it must be replaced by $\hat{\xi}_p$ in the estimate of $F'(\xi_p)/\bar{F}(\xi_p)$. Thus, the asymptotic variance of $n^{\frac{1}{2}}(\hat{\xi}_p - \xi_p)$ can be consistently estimated, enabling the construction of asymptotic confidence intervals for ξ_p .

We remark that for small samples the confidence intervals for ξ_p based on the above method are not appropriate. Several methods have been proposed for setting small sample confidence limits for ξ_p ; see Slud, Byar, and Green (1984), and the references cited therein.

3. WEAK CONVERGENCE OF THE QUANTILES OF THE KAPLAN-MEIER ESTIMATORS IN THE COMPETING RISKS PROBLEM.

In this section we use the representation $\hat{\xi}_j(p) = \xi_j(p) + (p - \hat{F}_j(\xi_j(p))) / F'_j(\xi_j(p)) + R_{j,n}(p)$ to obtain in a simple way the asymptotic distribution of $(\hat{\xi}_1(p), \dots, \hat{\xi}_m(p))$. In the sequel we use the notation $\hat{\xi}_j(p)$ rather than $\hat{\xi}_{j,p}$ when the quantiles of \hat{F}_j are viewed as a process; similarly for $\xi_j(p)$. Let G_j denote the distribution of $\min_{k \neq j} X_k$. The following proposition due to Aly et al. (1985) gives the strongest known result concerning the remainder process $R_{j,n}(p)$.

PROPOSITION 3.1. Suppose that the following conditions hold.

- (i) F_j and G_j are continuous.
- (ii) F_j is twice differentiable on (t_j, T_j) , where $t_j = \sup\{x: F_j(x)=0\}$ and $T_j = \inf\{x: F_j(x)=1\}$.
- (iii) $f_j(t) = F'_j(t)$ is positive on (t_j, T_j) .
- (iv) For some $p^* \in [0,1]$ we have

$$\sup_{0 \leq p \leq p^*} p |f'_j(\xi_j(p))| / (f_j(\xi_j(p))^2) < \infty$$

- (v) $0 < \lim_{t \rightarrow t_j^+} f_j(t) < \infty$.

Then, if $p_0 < p^*$ and $G_j(\xi_j(p_0)) < 1$, we have as $n \rightarrow \infty$

$$\sup_{0 \leq p \leq p_0} |R_{j,n}(p)| = O(n^{-\frac{3}{4}} (\log n)^{\frac{1}{2}} (\log \log n)^{\frac{1}{4}}) \text{ a.s.}$$

PROOF. See Corollary 4.1 of Aly et al. (1985).

For $s > 0$, let $D[0,s]$ be the space of all real valued functions defined on $[0,s]$ that are right continuous and have left limits, with the Skorohod metric topology, and let $D^m[0,s]$ denote the product metric space.

THEOREM 2. Suppose that the following conditions hold for each $j = 1, \dots, m$.

- (i) F_j is continuous.
- (ii) F_j is twice continuously differentiable and has a positive density on (t_j, T_j) .

(iii) For some $p^* \in [0,1]$ we have $\sup_{0 \leq p \leq p^*} \frac{p|f'_j(\xi_j(p))|}{(f_j(\xi_j(p)))^2} < \infty$.

(iv) $0 < \lim_{t \rightarrow t_j} f_j(t) < \infty$.

(v) $p_0 \in [0, p^*]$ is such that $\max_k F_k(\xi_j(p_0)) < 1$.

Then, as $n \rightarrow \infty$,

$$n^{\frac{1}{2}}(\hat{\xi}_1 - \xi_1, \dots, \hat{\xi}_m - \xi_m) \rightarrow (U_1, \dots, U_m)$$

weakly in $D^m[0, p_0]$, where U_1, \dots, U_m are independent mean 0 Gaussian processes with covariance structure given by

$$\text{Cov}(U_j(p_1), U_j(p_2)) = \frac{(1-p_1)(1-p_2)}{f_j(\xi_j(p_1)) f_j(\xi_j(p_2))} \int_0^{p_1} \frac{dw}{\prod_{k=1}^m F_k(\xi_j(w))(1-w)}$$

for $0 \leq p_1 \leq p_2 \leq p_0$.

PROOF. It follows from Theorem 10.1 of Aalen (1976) that

$$n^{\frac{1}{2}} \left(\frac{p - \hat{F}_1(\xi_1(p))}{f_1(\xi_1(p))}, \dots, \frac{p - \hat{F}_m(\xi_m(p))}{f_m(\xi_m(p))} \right) \rightarrow (U_1(p), \dots, U_m(p)) \text{ weakly in } D^m[0, p_0]. \quad (3.1)$$

The theorem follows from (3.1) and Proposition 3.1.

Theorem 2 is of statistical interest because it enables the construction of simultaneous confidence intervals for $\xi_1(p_1), \dots, \xi_m(p_m)$.

We remark that for fixed $p \in (0,1)$, if F_j' exists and is positive at ξ_j, p , and $\max_k F_k(\xi_j, p) < 1$ for $j = 1, \dots, m$, then as $n \rightarrow \infty$ $n^{\frac{1}{2}}(\hat{\xi}_{1,p} - \xi_{1,p}, \dots, \hat{\xi}_{m,p} - \xi_{m,p}) \rightarrow (U_1(p), \dots, U_m(p))$ in distribution. This is easily seen by substituting Theorem 1 for Proposition 3.1 in the proof of Theorem 2.

4. ESTIMATING THE QUANTILES OF A COHERENT SYSTEM.

Consider a coherent structure of m independent components. For the sake of concreteness, we keep in mind the specific example given by Figure 1 below.

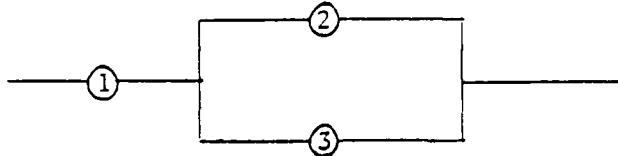


Figure 1. A coherent system of 3 components.

9

Let F denote the distribution function of the lifelength of the structure. Doss, Freitag, and Proschan (1986) (subsequently referred to as DFP) study the problem of estimating F under the following model. A sample of n systems, each with the same structure is available for testing. Each system is continuously observed until it fails. For every component in each system, either a failure time or a censoring time is recorded. A failure time is recorded if the component fails before or at the time of system failure. A censoring time is recorded if the component is still functioning at the time of system failure.

DFP propose the following procedure. Let F_1, \dots, F_m be the distribution functions of the lifelengths of the m components. Write

$$\bar{F}(t) = h(\bar{F}_1(t), \dots, \bar{F}_m(t)) \quad \text{for } t \geq 0, \quad (4.1)$$

where $h: [0,1]^m \rightarrow [0,1]$ is the reliability function (see Chapter 2 of Barlow and Proschan, 1981, for details concerning reliability functions). In the example given by Figure 1, it is easy to check that

$$\bar{F}(t) = \bar{F}_1(t)[1 - (1-\bar{F}_2(t))(1-\bar{F}_3(t))]$$

so that

$$h(u_1, u_2, u_3) = u_1[1 - (1-u_2)(1-u_3)] \quad \text{for } u_1, u_2, u_3 \in [0,1].$$

Let $\hat{\bar{F}}_1, \dots, \hat{\bar{F}}_m$ be the Kaplan-Meier estimates of $\bar{F}_1, \dots, \bar{F}_m$. To estimate \bar{F} , DFP propose

$$\hat{F}(t) = h(\hat{\bar{F}}_1(t), \dots, \hat{\bar{F}}_m(t)) \quad (4.2)$$

as an alternative to the naive estimate given by the proportion of systems still functioning at time t .

The purpose of this section is to study the estimates of quantiles

$$\hat{\xi}_j(p) = \hat{F}_j^{-1}(p) \text{ and } \hat{\xi}_p = \hat{F}^{-1}(p) \quad (0 < p < 1). \quad (4.1)$$

Before doing so we need to discuss the estimates \hat{F}_j and \hat{F} .

The weak convergence of the Kaplan-Meier estimator to a Gaussian process has been well-established in the literature (Breslow and Crowley, 1974; Aalen, 1976; Gill, 1983) under the assumption that the lifelengths and the censoring variables are independent. In our situation the component lifelengths are censored by the system lifelength, and the independence condition is clearly violated. We can, however, redefine the censoring variables to bypass this difficulty. This is easiest to explain in terms of the example given by Figure 1. Let X_j = lifelength of component j . Consider Component 1. Clearly, X_1 is censored by $Y_1 = X_2 \vee X_3$, which is independent of X_1 . Similarly, X_2 is censored by $Y_2 = X_1$, and X_3 by $Y_3 = X_1$. The construction is general: for an arbitrary system, X_j is censored by Y_j = lifelength of system if X_j is replaced by ∞ . DFP show that

(i) X_j and Y_j are independent.

(ii) The distribution function of Y_j is

$$G_j(t) = 1 - h(\bar{F}_1(t), \dots, \bar{F}_{j-1}(t), 1, \bar{F}_{j+1}(t), \dots, \bar{F}_m(t)). \quad (4.4)$$

The main result of DFP can now be stated.

PROPOSITION 4.1. Suppose that F_1, \dots, F_m are continuous, and let T satisfy $\max_{1 \leq j \leq m} F_j(T) < 1$.

Then as $n \rightarrow \infty$

$$n^{1/2}(\hat{F}_1 - F_1, \hat{F}_2 - F_2, \dots, \hat{F}_m - F_m) \xrightarrow{D} (W_1, W_2, \dots, W_m)$$

weakly in $D^m[0, T]$, where W_1, \dots, W_m are independent mean 0 Gaussian processes. The covariance structure of W_j is given by

$$(i) \quad \text{Cov}(W_j(t_1), W_j(t_2)) = \bar{F}_j(t_1)\bar{F}_j(t_2) \int_0^{t_1} \frac{dF_j(u)}{\bar{G}_j(u)(\bar{F}_j(u))^2} \text{ for } 0 \leq t_1 \leq t_2 \leq T.$$

$$(ii) \quad n^{1/2}(\hat{F} - F) \xrightarrow{D} W \text{ weakly in } D[0, T],$$

where W is a mean 0 Gaussian process with covariance structure given by

$$\text{Cov}(W(t_1), W(t_2)) = \sum_{j=1}^m h_j(t_1)h_j(t_2) \bar{F}_j(t_1)\bar{F}_j(t_2) \int_0^{t_1} \frac{dF_j(u)}{\bar{G}_j(u)(\bar{F}_j(u))^2}$$

where

$$h_j(t) = \frac{\partial h}{\partial u_j}(u_1, \dots, u_m) \Big|_{(u_1, \dots, u_m) = (\bar{F}_1(t), \dots, \bar{F}_m(t))}. \quad (4.5)$$

The model being considered is a generalization of the competing risks model, which corresponds to a series system of m components. The following result generalizes Theorem 2 to this model.

PROPOSITION 4.2. Let $\hat{\xi}_j(p)$ be defined by (4.3), and suppose that the conditions of Theorem 2 are satisfied. Then as $n \rightarrow \infty$

$$n^{\frac{1}{2}}(\hat{\xi}_1 - \xi_1, \dots, \hat{\xi}_m - \xi_m) \xrightarrow{D} (U_1, \dots, U_m)$$

weakly in $D^m[0, p_0]$, where U_1, \dots, U_m are independent mean zero Gaussian processes.

The covariance structure of U_j is given by

$$\text{Cov}(U_j(p_1), U_j(p_2)) =$$

$$\frac{(1-p_1)(1-p_2)}{f_j(\xi_j(p_1)) f_j(\xi_j(p_2))} \int_0^{p_1} \frac{dw}{\bar{G}_j(\xi_j(w))(1-w)^2} \quad \text{for } 0 \leq p_1 \leq p_2 \leq p_0,$$

where G_j is given by (4.4).

PROOF. Substitute Part (i) of Proposition 4.1 for Theorem 10.1 of Aalen (1976) in the proof of Theorem 2. ■

The main result of this section gives a Bahadur representation for $\hat{\xi}_p$ analogous to (1.5), and is useful in obtaining the asymptotic distribution of $\hat{\xi}_p$.

THEOREM 3. Let $0 < p < 1$ and assume the following conditions.

(i) F_1, \dots, F_m are continuous, with $\max_{1 \leq j \leq m} F_j(\xi_p) < 1$.

(ii) For each j , F_j has a positive derivative at ξ_p .

Then,

$$\hat{\xi}_p = \xi_p + \frac{p - \hat{F}(\xi_p)}{F'(\xi_p)} + R_{n,p},$$

where $n^{\frac{1}{2}} R_{n,p} \rightarrow 0$ in probability as $n \rightarrow \infty$.

PROOF. The proof we give is a straightforward modification of the proof of Theorem 1.

Since h is strictly increasing in each of its arguments (see Theorem 1.2 of Barlow and Proschan, 1981) it follows that ξ_p is the unique number satisfying $F(\xi_p) = p$. Furthermore, h is differentiable over $[0,1]^m$ (see Lemma 2.1 of DFP), so by the chain rule $F'(\xi_p)$ exists and is given by

$$F'(\xi_p) = \sum_{j=1}^m h_j(\xi_p) F'_j(\xi_p)$$

where h_j is given by (4.5).

For any $t \geq 0$ and any n define

$$Z'_{t,n} = n^{\frac{1}{2}} \left[F\left(\xi_p + \frac{t}{n^{\frac{1}{2}}}\right) - \hat{F}\left(\xi_p + \frac{t}{n^{\frac{1}{2}}}\right) \right] / F'(\xi_p) \quad \text{and}$$

$$W'_n = n^{\frac{1}{2}}(p - \hat{F}(\xi_p)) / F'(\xi_p).$$

We have $\{n^{\frac{1}{2}}(\hat{\xi}_p - \xi_p) \leq t\} = \{Z'_{t,n} \leq t_n\}$, where $t_n = n^{\frac{1}{2}} \left[F\left(\xi_p + \frac{t}{n^{\frac{1}{2}}}\right) - p \right] / F'(\xi_p) \rightarrow t$

(cf. equations (2.9) and (2.10)). As in the proof of Theorem 1, it suffices to show that $Z'_{t,n} - W'_n \rightarrow 0$ in probability.

Using (4.2) we write

$$Z'_{t,n} - W'_n = \frac{n^{\frac{1}{2}}}{F'(\xi_p)} \left[\left[h\left(\hat{F}\left(\xi_p + \frac{t}{n^{\frac{1}{2}}}\right)\right) - h\left(\hat{F}(\xi_p)\right) \right] - \left[h\left(\bar{F}\left(\xi_p + \frac{t}{n^{\frac{1}{2}}}\right)\right) - h\left(\bar{F}(\xi_p)\right) \right] \right], \quad (4.6)$$

where for economy of notation $\hat{\underline{F}}(s)$ and $\bar{\underline{F}}(s)$ denote the vectors $(\hat{F}_1(s), \dots, \hat{F}_m(s))$ and $(\bar{F}_1(s), \dots, \bar{F}_m(s))$, respectively. We now apply the Mean Value Theorem (see for example Apostol, 1964, Theorem 6.17) to each of the differences inside the brackets in (4.6): there exist points \hat{a} and a lying on the line segments joining $\hat{\underline{F}}(\xi_p + \frac{t}{n^{\frac{1}{2}}})$ and $\hat{\underline{F}}(\xi_p)$, and $\bar{\underline{F}}(\xi_p + \frac{t}{n^{\frac{1}{2}}})$ and $\bar{\underline{F}}(\xi_p)$, respectively, such that the right side of (4.6) is equal to

$$\frac{n^{\frac{1}{2}}}{F'(\xi_p)} \left\{ \nabla h(\hat{a}) \cdot \left[\hat{\underline{F}}(\xi_p + \frac{t}{n^{\frac{1}{2}}}) - \hat{\underline{F}}(\xi_p) \right] - \nabla h(a) \cdot \left[\bar{\underline{F}}(\xi_p + \frac{t}{n^{\frac{1}{2}}}) - \bar{\underline{F}}(\xi_p) \right] \right\}, \quad (4.7)$$

where ∇h denotes the gradient of h . It is convenient to rewrite (4.7) as

$$\begin{aligned} & \left\{ \frac{n^{\frac{1}{2}}}{F'(\xi_p)} \nabla h(\hat{a}) \cdot \left[\left[\hat{\underline{F}}(\xi_p + \frac{t}{n^{\frac{1}{2}}}) - \hat{\underline{F}}(\xi_p) \right] - \left[\bar{\underline{F}}(\xi_p + \frac{t}{n^{\frac{1}{2}}}) - \bar{\underline{F}}(\xi_p) \right] \right] \right\} \\ & + \left\{ \frac{n^{\frac{1}{2}}}{F'(\xi_p)} (\nabla h(\hat{a}) - \nabla h(a)) \left[\bar{\underline{F}}(\xi_p + \frac{t}{n^{\frac{1}{2}}}) - \bar{\underline{F}}(\xi_p) \right] \right\}. \end{aligned} \quad (4.8)$$

Since we clearly have $\hat{a} \rightarrow \bar{\underline{F}}(\xi_p)$ a.s. and $a \rightarrow \bar{\underline{F}}(\xi_p)$, the continuity of ∇h (see Lemma 2.1 of DFP) and the differentiability of each F_j at ξ_p imply that the second term inside the braces in (4.8) converges to 0 in probability. The first term inside the braces in (4.8) converges to 0 in probability since by (2.11), for each j

$$n^{\frac{1}{2}} \left\{ \left[\hat{F}_j(\xi_p + \frac{t}{n^{\frac{1}{2}}}) - \hat{F}_j(\xi_p) \right] - \left[\bar{F}_j(\xi_p + \frac{t}{n^{\frac{1}{2}}}) - \bar{F}_j(\xi_p + \frac{t}{n^{\frac{1}{2}}}) \right] \right\} \rightarrow 0 \text{ in probability,}$$

and the components of ∇h are bounded by 1 (see Lemma 2.1 of DFP). This completes the proof of Theorem 3.

Theorem 3 immediately implies the asymptotic normality of $\hat{\xi}_p$.

COROLLARY 4.1. Under the conditions of Theorem 3, $n^{\frac{1}{2}}(\hat{\xi}_p - \xi_p) \xrightarrow{d} N(0, \sigma^2(p))$, where

$$\sigma^2(p) = \frac{1}{F'(x_p)} \sum_{j=1}^m \left[h_j(x_p) \bar{F}_j(x_p) \right]^2 \int_0^{x_p} \frac{dF_j(u)}{\bar{G}_j(u)(\bar{F}_j(u))^2} .$$

To construct asymptotic confidence intervals for ξ_p we need to be able to consistently estimate the quantity $\sigma^2(p)$. An extension of the argument given at the end of Section 2 shows that this can be done under the additional assumption that for each j , F_j is continuously differentiable in a neighborhood of x_p . Details are omitted.

REFERENCES

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. Scand. J. Statist. 3, 15-27.

Aalen, O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. Scand. J. Statist. 5, 141-150.

Aly, E.E., Csörgő, M., and Horváth, L. (1985). Strong approximations of the quantile process of the product-limit estimator. J. Multivar. Anal. 16, 185-210.

Anderson, P.K. and Borgan, O. (1984). Counting process models for life history data: a review. Tenth Nordic Conf. in Math. Stat. To appear in Scand. J. Statist. .

Apostol, T.M. (1964). Mathematical Analysis. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts.

Barlow, R.E. and Proschan, F. (1981). Statistical Theory of Reliability and Life Testing. To Begin With, Silver Springs, Maryland.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimators under random censorship. Ann. Statist. 2, 437-453.

Doss, H., Freitag, S., and Proschan, F. (1986). Estimating jointly system and component life distributions using a mutual censorship approach. Technical Report No. M717, Department of Statistics, Florida State University.

Földes, A., Rejtő, L. and Winter, B.B. (1980). Strong consistency properties of nonparametric estimators for randomly censored data (Part I). Periodica Math. Hungar. 11, 233-250.

Gail, M. (1975). A review and critique of some models used in competing risks analysis. Biometrics 31, 209-222.

Ghosh, J.K. (1971). A new proof of the Bahadur representation of quantiles and an application. Ann. Math. Statist. 42, 1957-1961.

Gill, R.D. (1980). Censoring and Stochastic Integrals. Mathematical Centre Tracts 124. Mathematisch Centrum, Amsterdam.

Gill, R.D. (1983). Large sample behaviour of the product-limit estimator on the whole line. Ann. Statist. 11, 49-58.

Hall, W.J. and Weilner, J.A. (1980). Confidence bands for a survival curve from censored data. Biometrika 67, 133-143.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

Lenglart, E. (1977). Relation de domination entre deux processus. Ann. Inst. Henri Poincaré 13, 171-179.

Lipster, R.S. and Shirayev, A.N. (1978). Statistics of Random Processes II, Springer-Verlag, New York-Heidelberg-Berlin.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. Ann. Statist. 11, 453-466.

Slud, E.V., Byar, D.P., and Green, S.B. (1984). A comparison of reflected versus test-based confidence intervals for the median survival time, based on censored data. Biometrics 40, 587-600.

Wellner, J. (1985). A heavy censoring limit theorem for the product limit estimator. Ann. Statist. 13, 150-162.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 382	2. GOVT ACCESSION NO. AD-A175050	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Quantiles of Kaplan-Meier Estimator	5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT	
7. AUTHOR(s) Hani Doss and Steven Freitag	6. PERFORMING ORG. REPORT NUMBER N00014-86-K-0156	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267	
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Code 1111	12. REPORT DATE November 4, 1986	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	13. NUMBER OF PAGES 19	
	15. SECURITY CLASS. (of this report) UNCLASSIFIED	
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Also issued as Florida State University Statistics Report No. M742 prepared under Air Force Office of Scientific Research No. 86-195.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Quantiles; competing risks problem; Bahadur representation; Kaplan-Meier estimator; coherent structure; reliability function; stochastic integral.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) We use the theory of counting processes, martingales and stochastic integration, to establish in a simple way a Bahadur representation for the quantiles of the Kaplan-Meier estimator. We combine this Bahadur representation with a result of Aalen (1976) to prove the asymptotic independence of the quantile estimates in the competing risks problem. Finally, we use the theory to study the estimates of the quantiles of the life-length of a coherent system proposed by Doss, Freitag, and Proschan (1986).		

E

W

D

87

DT / C